Stoch Environ Res Ris Assess
DOI 10.1007/s00477-007-0113-6

ORIGINAL PAPER

A simple method for effective multi-site generation of stochastic hydrologic time series

Nesa Ilich · Jovan Despotovic

Abstract This paper presents an algorithm for generating stationary stochastic hydrologic time series at multiple sites. The ideas in this paper constitute a radical departure from commonly accepted methodologies. The approach relies on the recent advances in statistical science for simulating random variables with arbitrary marginal distributions and a given covariance structure, and on an algorithm for re-ordering the generated sub-sets of each synthetic year of data such that the annual auto-correlation of desired lag is maintained, along with the autocorrelations between the end of the preceding year and the beginning of the current year. The main features of the proposed algorithm are simplicity and ease of implementation. A numerical test is presented containing the generation of 1000 years of weekly stochastic series for four sites based on the 84 years of historical natural weekly flows from Southern Alberta in Canada.

1 Introduction

The early efforts to generate hydrologic time series date back to the 1960s (Thomas and Fiering 1962; Matalas 1967). Although originally motivated by possible applications in finance, the work of Box and Jenkins (1970) on time series analyses has had a profound impact on researchers in stochastic hydrology over the last three decades. There are a number of variants of the approach of Box and Jenkins in stochastic hydrology promoting the auto-regressive moving average models in various forms, most frequently in combination with disaggregation models which require that the annual series be generated first, to ensure the annual statistics are preserved, and then be broken into seasonal (typical monthly) time steps using various disaggregation algorithms (Valencia and Schaake 1973; Mejia and Rousselle 1976). The majority of publications dealing with the synthetic generation of hydrologic time series that were published since the mid 1970s are in some ways related to the initial work of Box and Jenkins (1970). A good review of the history of previous efforts is provided by Srinivas and Srinivasan (2005).

There is currently no universally accepted methodology nor a readily available user-friendly computer program that has gained widespread acceptance among stochastic hydrologists. The reasons may be many, and they are certainly in part due to the complex nature of hydrologic processes, including discontinuities in the data in dry regions, the controversy regarding the Hurst phenomenon (Hurst 1957), or the uncertainty proposed by some researchers regarding the assumption that hydrologic phenomena are indeed stationary in the long term, prompting the efforts to include the concept of a ‘‘sudden shift’’ pattern into modeling (Sveinsson and Salas 2003). The current lack of universal approach may in part be due to the complexity of the methods rendered so far, which involve significant effort and knowledge to conduct identification of the appropriate model and estimation of its parameters, as well as the difficulties in assessing the shape of the multivariate probabilities and their transformations from

N. Ilich
University of Calgary, Calgary, AB, Canada

N. Ilich (✉)
7128 - 5 Street NW, Calgary, AB, Canada T2K 1C8
e-mail: nilich@optimal-solutions-ltd.com

J. Despotovic
Department of Civil Engineering,
University of Belgrade, Belgrade, Serbia
e-mail: edespoto@hikom.grf.bg.ac.yu

normal to skewed distribution required to model hydrologic processes. The gap between theory and practice has haunted stochastic hydrologists to this day. Stochastic modeling of seasonal flows has proven to be difficult due to periodic nature of stochastic parameters coupled with the recognized difficulties of modeling auto-regressive behavior of higher orders, thus restricting most efforts to date mainly to AR(1) monthly modeling. As Srinivas and Srinivasan (2005) point out, in spite of the numerous reports on modeling efforts in stochastic hydrology, none have gained universal acceptance. In fact, some agencies such as the United States Bureau of Reclamation (USBR) resort instead to a simplistic approach of recycling subsets of historical flow series, an approach known as the index sequential sampling (ISM) method (Kendall and Dracup 1991) rather than rely on any of the complex models which had been the subject of so much research efforts in the past few decades. In addition to the need to handle multi-site generation, there is also a pressing need in the industry for modeling of flows based on shorter (weekly or daily) time steps as attempted by some researchers (Rasmmussen et al. 1996; Aksoy and Bayazit 2000). The work in this paper borrows ideas from recent developments in statistical science applicable to modeling of random variables with given marginal distribution functions and a given covariance structure, and extends these developments to suit the needs of modeling hydrologic time series. Since this approach radically differs from other published research efforts in stochastic hydrology, it is prudent to first clearly state the assumptions and the ultimate objectives prior to formulating the algorithm, which is done in Sect. 2. Section 3 explains data representation and gives hints on the basic ideas of the algorithm, which are explained in more detail in Sect. 4. Section 5 shows the results of a numerical example conducted on four locations of interest in southern Alberta, Canada, each complete with 84 years of historical naturalized weekly flows. This is followed by closing remarks, acknowledgements and references.

2 Model assumptions

The ideas in this paper are currently restricted to modeling continuous and stationary time series of weekly or monthly flows at multiple sites. Stochastic hydrologic time series should have similar statistical properties as the historical series. While it is recognized that hydrologic time series represent continuous natural processes, they are invariably represented as discrete series of average values over selected (typically monthly or weekly) time steps. As such, they are described using statistics such as the means, standard deviations, skews and probability distributions. In addition, flows are usually auto-correlated as well as cross-correlated when multiple sites are considered. Of particular interest to annual series so far have been higher order auto-correlations due to the assumed existence of long term trends affected by phenomena which are not completely understood (but are often seen in the data), such as for example the El Nino effect. Successful generation of stochastic hydrologic time series in this paper is based on the following assumptions:

(a) Partial historical series containing only the data for a given month or week has a unique statistical distribution that should be matched in the simulated series. This distribution function can be represented either by one of the known theoretical distributions that fits the data well, or by using a new generation of kernel-based empirical distributions (Lall 1995).
(b) Statistics such as the historical mean and standard deviation of each month (or week) should be preserved in the stochastic series;
(c) Historical auto-correlations and cross-correlations should match the simulated;
(d) Annual statistics of the historical series, such as the annual mean, standard deviation, auto-correlations and cross-correlations should also match the annual statistics of the simulated series.

As mentioned above, the model presented here is stationary, e.g. there is no attempt to model phenomena such as sudden shifts in the statistical properties of the generated series. It is commonly accepted that a randomly generated series that well preserves the above listed set of historical statistics is a good example of a stochastic hydrologic time series. If it can be agreed that the above statistics summarize the desired properties of a stochastic hydrologic series, the objective of this paper is to demonstrate that there is an easy and computationally efficient way to achieve this. Recall that most of the previous methods attempted to generate synthetic stream flows by using the linear auto-regressive model in the form of:

$$y_t = a_{t,t-1}y_{t-1} + a_{t,t-2}y_{t-2} + \cdots + a_{t,0}y_0 + \varepsilon_t \quad (1)$$

where stream flows at time step t are linked to the flows at previous time steps $t-1$, $t-2$, etc., through a linear regression model with parameters $a_{t,\,t-i}$ that form a lower triangular matrix of regression coefficients for sequential values of the time index $t = 1, 2, \ldots, n$. Each set of coefficients corresponds to regression coefficients of flow in a particular time step t treated as a variable statistically dependent on the flows in the previous time steps. Hence, the first flow value can be determined in a random manner, but all other subsequent flows exhibit statistical dependence to the previously generated value(s) by employing the model defined by Eq. 1. Such models would then have to generate synthetic flows in sequential order. One of the principal problems with this model was its inability to preserve the skewed statistical distribution of flows. This is due to the fact that the random term ε in Eq. 1 is normally distributed, and as a result of that the generated flows eventually end up with normal distribution. Log transformations of the original data have been suggested and incorporated in some models such as the HEC-4 (US Army Corps of Engineers 1971), but with success limited only to average monthly flows. A standard widely accepted tool for generating weekly or daily flows for multiple sites that preserves all relevant statistical dependence along with the probability distribution function has yet to be established.

It is known from statistical science that statistical dependence defined by the multiple regression model (Eq. 1) is equivalent to the statistical dependence between variables y_t, $y_{t\text{-}1}$, $y_{t\text{-}2}$, ..., y_0 and it could have been represented by a Pearson correlation matrix. In fact, algorithms that can estimate the lower triangular matrix of regression coefficients $a_{t,t-i}$ as a function of the available product moment correlation matrix have been available for some time (Cooley and Lohnes 1971), and a computer programs that can provide this transformation are available (UNESCO 2006). It can therefore be established that fitting random variables y_t in model (Eq. 1) is equivalent to generating random variables y_t which have the same mutual correlation structure. The principal idea in this paper is to rely on the available algorithms from statistical science to generate random variables with arbitrary marginal distribution and given covariance structure. To the best of our knowledge, these algorithms have not yet been used in stochastic hydrology. They allow both preservation of the original statistical properties of hydrologic data and their statistical dependence.

3 Data representation and basic ideas

Without any loss of generality, simulated weekly naturalized flows are assumed as an example of a stochastic hydrologic series throughout the rest of this paper. Consider a historical data organization of multiple stations in a matrix format as depicted in Table 1. Historical data can be viewed as a matrix containing each year of data in a single row, while each column represents data for a particular week, starting from week 1 to week 52 for a single station. If three stations are considered simultaneously, the data matrix would have a total of 156 columns while the number of rows equals the number of available years of data.

The proposed algorithm for stochastic generation requires historical input data in the form of the above matrix populated by weekly naturalized flows for all stations under consideration. The matrix should contain as many years of available data as possible, and the occasional missing values should be filled using one of the common techniques. Once the historical database is complete, it is possible to obtain the necessary statistics of the historical series that the proposed algorithm should be able to replicate in the simulated series, such as the weekly means, standard deviation, skew, probability distribution for a given week and the upper triangular matrix of correlation coefficients $\sigma_{i,j}$ shown in Table 2 which describe statistical

Table 1 Matrix representation of historical data

Year	Station 1					Station 2					Station 3				
	Weeks					Weeks					Weeks				
	1	2	.	.	52	1	2	.	.	52	1	2	.	.	52
1933	$X_{1,1}$	$X_{1,2}$	.	.	$X_{1,52}$	$X_{1,53}$	$X_{1,54}$	.	.	$X_{1,104}$	.	.	.	.	$X_{1,156}$
1934	.														.
1935	.														.
1936	.														.
1937	.														.
1938	.		.	.	.	.	.	$X_{i,j}$	.	.	.	.	.	.	.
.	.														.
.	.														.
.	.														.
1999	.														.
2000	.														.
2001	.														.
2002	.	.	.	.	.	.	.	.	.	.	.	.	.	.	$X_{m,n}$

Table 2 Schematic layout of correlation matrix

Variable	Station 1					Station 2				Station 3				
	Variable					Variable				Variable				
	1	2	3	.	52	53	54		104	105	106	.	155	156
1	1	$\sigma_{1,2}$	$\sigma_{1,3}$		$\sigma_{1,52}$	$\sigma_{1,53}$	$\sigma_{1,54}$		$\sigma_{1,104}$	$\sigma_{1,105}$	$\sigma_{1,106}$			
2		1	$\sigma_{2,3}$		$\sigma_{2,52}$	$\sigma_{2,53}$	$\sigma_{2,54}$		$\sigma_{2,104}$	$\sigma_{2,105}$	$\sigma_{2,106}$			
3			1											
4				1										
5					1									
6						1								
.							1							
.								.						
.								.						
.								.						
.								.						
155													1	
156														1

dependence between any two columns of the matrix in Table 1. It is worthwhile to examine the nature and significance of information contained in the correlation matrix, hence its structure is schematically depicted in Table 2. Flows in each week are denoted as a variable with index i for rows and j for columns. The diagonal correlation coefficients are set to 1, as they represent correlation of the flow in a given week to itself. Coefficient $\sigma_{1,2}$ describes statistical dependence of the flows in the first and second week of January, since it was obtained by calculating correlation between the first two columns in the matrix shown in Table 1. Similarly, coefficients $\sigma_{2,3}$ and $\sigma_{1,3}$ describe statistical dependence between the flows in week 3 and the flows in week 2 and 1, respectively. In essence, these coefficients are a measure of autocorrelation of weekly flows.

Similarly, cross correlation between different stations is also contained in the correlation matrix. For example, coefficients $\sigma_{1,53}$, $\sigma_{1,105}$ and $\sigma_{53,105}$ represent cross correlations between flows in the first week among stations [1,2], [1,3] and [2,3], respectively. Since the correlation matrix maps the flow at each site and in each week to all other weekly flows at all sites, it also includes information about possible lagged cross-correlation between different sites. Two or more stations located in a given river basin may have a lag related to the transient flow propagation lasting more than a week, such that the peak flow in week i–1 at an upstream station shows in week i at a downstream station. This is a form of lagged cross-correlation. Such statistical dependence is also worth preserving in the simulated data, although very little has been documented in the literature on the efforts to achieve this. Most existing algorithms consider cross-correlation as statistical dependence that is worth preserving only within simultaneous time steps. Given the above data representation, the obvious question that comes to mind is ‘‘can the weekly flows stored in matrix columns be replaced with their random equivalents, such that their historical statistical distributions, means, standard deviations and skews be preserved, while simultaneously preserving the correlation matrix derived from the historical data’’? Recent developments in statistical science have resulted in algorithms that can achieve this. Such algorithms were designed to generate random variables with arbitrary marginal distributions and a given covariance structure, such that a matrix similar to the one shown on Table 1 but with 1,000 rows (to represent 1,000 simulated years) can be populated with random weekly flows that conform to the historical weekly statistics, including the distribution function, and conforming to the given covariance structure obtained from historical flows. One could then introduce an algorithm for re-ordering the rows of this matrix such that a configuration (or a sequence) of rows is found that provides a good match for the remaining statistics, namely the auto-correlation from the last weeks of a given year to the first weeks of the subsequent year, as well as the annual auto-correlation for all stations. For a randomly generated matrix of 1,000 rows, there are $1{,}000^2$ available combinations of how the rows could be re-ordered. The principal goal of this paper is to present one such algorithm for re-ordering of the rows of such a matrix that achieves the stated goals. Note that re-ordering of the rows of this matrix in no way affects its established correlation structure, for the same reasons that the correlation matrix in Table 2 remains unchanged if the rows of the matrix in Table 1 undergo permutation. This, in a nutshell, is the basic idea behind the proposed algorithm.

4 A three-step process

The three distinct steps are: (1) generated weekly stochastic flows independently for each week; (2) reorder the flows to induce desired correlation within a year; and (3) reorder the years of generated data to induce desired year to year correlations.

4.1 Step 1. Generate weekly stochastic flows

It is known that the statistical distribution of flows may be different throughout the year due to different runoff and climatic conditions. Hence, the first step is to generate randomly 1,000 years of weekly stochastic flows that would have the desired statistical properties. This can be done by treating the data for each week as an independent sample and fitting a statistical distribution to it using the maximum likelihood approach, and then running a Monte Carlo simulation for the chosen distribution functions and its estimated parameters. Much has been written about fitting statistical distributions in hydrology and there is no reason to repeat it here. In addition to using theoretical distribution functions known to work well in hydrology, a new nonparametric approach (Lall 1995) that has been emerging lately holds out a promise to overcome some of the earlier problems associated with fitting theoretical functions to historical data samples. As Sharma et al. (1997) explain, the so called Kernel density functions are de facto a weighted moving average of the empirical frequency distribution of the available data. The result is a distribution function guaranteed to fit in the historical data in the probability range for which the data are available, which virtually eliminates the need to run the goodness of fit tests required for fitting theoretical distributions. Much of the on-going research in nonparametric density functions has been focused on handling of their tail ends. Lall (1995) provides a summary of several approaches studied so far, of which the approach by Moon et al. (1993) has been used in the numerical examples in this paper.

The advantage of independent Monte Carlo generation of stochastic flows in each week is further enhanced by a monitoring technique which keeps track of the desired statistics of the generated series, and terminates the generation when these are considered sufficiently close to the historical statistics, using a prescribed tolerance limit as a target. This technique is not essential for working of the proposed method, but it does provide an additional control mechanism for fine-tuning the fit between the historical and generated statistics. Instead of generating only 1,000 weekly synthetic flows, the algorithm could continue to generate flows and recalculate the statistics of the 1,000 elements of the series. Thus, after a random generation of 1,001st flow, the statistics such as the mean and standard deviation can be calculated for a series [2, 1,001] and compared with the initial series [1, 1,000]. This comparison continues for subsequent generated flows for a given week, introducing a form of 'tournament' selection among them, which ends when the simulated statistics are very close to historical, or when a specified number (e.g. 10,000) of generated variables has been reached, in which case the set of 1,000 elements with the statistics closest to the historical targets is selected. Note that recalculation of statistics such as the means and standard deviations does not involve recalculating the entire sums, but only updating them by removing the first element in a 1,000 years set and adding a newly generated element to it. Statistics of the resulting synthetic flows can thus match their historical targets with any desired precision, as demonstrated in Sect. 5 below.

4.2 Step 2. Reorder weekly flows to construct desired correlations

It is obvious that Step 1 simulates weekly flows which are independent of each other, since the only input in this step is the statistical distribution function of flows for each week. Step 1 provides desirable properties of weekly flows in terms of their statistical distribution, mean and standard deviation, however it does not address statistical dependence among flows in subsequent weeks, since correlation among independent random variables is zero. To address this, Step 2 takes the flows generated in Step 1 and reorders them such that the desired correlation (statistical dependence between the flows in subsequent weeks) is induced. In practical terms, this means that flow in week 2 from year n may be swapped with the flow in week 2 for year m, if such a swap increases the overall correlation between weeks 1 and 2, where n and m can be any two years from a simulated sample.

Step 2 results in permutations of the data in each column of the matrix of random weekly flows generated in Step 1 until the desired covariance structure is induced. The desired covariance structure is obtained from the historical naturalized flow series, as already mentioned in Sect. 3. There are known algorithms that can be used to achieve this goal, most notably the one proposed by Iman and Conover (1982). This algorithm first creates a matrix of uniform random variables with a desired covariance structure, and then uses it as a key for re-ordering the elements in the matrix of generated random variables with a desired marginal distribution. Other variants of this idea have been published (Devore 1986; Cario and Nelson 1996, 1997; Ghosh and Henderson 2002), however the work of Iman and Conover (1982) remains pivotal as a basis for most commercial applications in this field so far. At the end of Step 2, the original matrix generated in step 1 has been transformed such that it conforms to the desired covariance structure. However, the transformation involved only a permutation of the elements of matrix within each column. By doing so, the initial achievements of step 1 related to the preservation of the statistical distribution functions, means, and standard deviations of the generated flows for each week have not been violated.

Although the algorithm of Iman and Conover (1982) proposed in step 2 requires the rank-order correlation matrix as its input (rather than the product–moment correlation that is traditionally of interest in hydrology), and matches it with the rank order correlation of the permuted matrix of randomly generated series, this usually also results in a close enough match (by association) of the product moment correlations between the historical and simulated data. The downside of this algorithm is that if the match is less than perfect, there are no fine-tuning parameters to improve it. Subsequently, successful experiments have been conducted with variations on this theme by using the lower triangular matrix of regression coefficients as input instead of the covariance matrix. The matrix of regression coefficients can be obtained from the matrix of product moment correlation coefficients (Cooley and Lohnes 1971). In this approach, it is possible to control the accuracy of the reordering process by introducing a step-wise progression that allows adjustable convergence factors. The proposed alternative algorithm can also be used for inducing statistical dependence among random variables of arbitrary marginal distributions. It is not covered here as it is a topic of interest to operations research practitioners, and it is a subject matter of a separate paper (Ilich 2006). The algorithm is essentially a variant of the one proposed by Iman and Conover (1982) since the basic idea of reordering the existing elements of random matrix remains the same, only the technique of doing it is somewhat modified. It should be mentioned however that the accuracy of the algorithm of Iman and Conover (1982) obtained in tests conducted so far has been acceptable for most practical intents and purposes.

At the end of this step, not only have the weekly statistics within each year and each station been matched, but the annual means, standard deviations and cross-correlation have also automatically been brought to compliance with the desired targets by their association with the weekly flows, as demonstrated by the numerical example in Sect. 5.

4.3 Step 3. Reorder simulated years to conform to the remaining statistics

So far, the treatment of flows has had nothing to do with a classical notion of 'time series' as defined in the work of Box and Jenkins (1970) and other authors, since the entire process so far was conducted based on the ability to generate random variables with desired frequency distribution and covariance structure. In time series, however, the auto-regressive relationships continue from year to year, both on an annual and on a weekly basis (e.g. flows in weeks 50, 51, 52 in any given year may exhibit statistical relationships to flows in weeks 1, 2, and 3 etc. of the subsequent year). Also, annual auto-correlations should be preserved for all stations in the generated series. Step 2 addressed neither the weekly correlations across the year end, nor did it address the desired annual auto correlation structure. Of significant importance to this step is a realization that permutation of rows of data in a matrix where each row represents a hypothetical annual series of weekly (or monthly) data does not alter the covariance structure of this matrix. This practically means that permutation of the entire rows of such a matrix would not jeopardize the results of the previous Step 2 of the proposed algorithm, which is focused on inducing a correlation structure of flows within a year.

At this point it is necessary to introduce correlation among flows in weeks that cross from year to year. Let n represent an index of week 52 for a given station and $\sigma_{n,1}$, the correlation coefficient of historical data in Table 1 between flows in week 52 of any given year and flows in week 1 of the subsequent year (hereinafter ' is related to the data in the subsequent year). Similarly, $\sigma_{n\text{-}1,1}$, relates flows in week 51 to the flows in the first week of the following year, and $\sigma_{n,2}$, relates flows in week 52 to the flows in the second week of the subsequent year. The same notation is applicable to correlations $\rho_{n,1}$, $\rho_{n-1,1}$, and $\rho_{n,2}$, of simulated flows. To replicate an equivalent of an auto-regressive model of lag 2 (i.e. AR(2) model) for weeks that cross from year to year, consider the following sum of squares error function D_k, which is a sum of deviations of simulated correlations from their historical targets for a given station k:

$$D_k = (\rho_{n-1,1'} - \sigma_{n-1,1'})^2 + (\rho_{n,1'} - \sigma_{n,1'})^2 + (\rho_{n,2'} - \sigma_{n,2'})^2 \quad (2)$$

For a general lag L the equivalent AR(L) representation function D_k is defined as

$$Dk = \sum_{p=1}^{L} \sum_{q=n-L+p}^{n} (\rho_{q,p}^{k} - \sigma_{q,p}^{k})^2 \tag{3}$$

where q and p represent indices of the weeks in the current and subsequent years, respectively, while n is the total number of time steps in a year (for weekly time steps $n = 52$). With a sufficiently large synthetic series, it is possible to find an order of the rows in the simulated matrix that brings the value of D_k sufficiently close to zero, such that it can be argued that the simulated weekly flows at station k exhibit the desired statistical dependence observed in the historical time series. Similarly, denote with AH_l and AS_l simulated and historical annual auto-correlation function of lag l. Revise the above equation for D_k by adding the terms for matching the historical and simulated annual auto-correlations:

$$Dk = \sum_{p=1}^{L} \sum_{n=n-L+p}^{n} (\rho_{q,p}^{k} - \sigma_{q,p}^{k})^2 + \sum_{l=1}^{m} (AH_l^k - AS_l^k)^2 \tag{4}$$

where m refers to the number of years for which annual auto correlation in the historical series should be similar to the annual auto correlation of the synthetic series. Equation 4 represents the squared sum of errors in reproducing the observed auto correlations across the year end and the annual auto-correlations of the simulated series. Re-ordering the rows of the simulated matrix such that the value of the right hand side of Eq. 4 is sufficiently close to zero is equivalent to matching both the desired weekly and annual statistics for station k. To extend the above observations to multiple stations, introduce a composite distance function D as a sum of all individual functions D_k:

$$D = \sum_{k=1}^{r} D_k \tag{5}$$

where r is the total number of stations being considered simultaneously. A sufficiently small tolerance limit ε can be introduced for D, such that an appropriate order of the rows of simulated matrix is found when $D < \varepsilon$. The search for the appropriate order is conducted using a combinatorial algorithm. For a sufficiently large sample, there are many possible sequences of rows that will meet the desired criteria, while it is sufficient to find only one. One important aspect of the rows re-ordering algorithm is that it should maintain relative uniformity of the statistical properties throughout the simulated series. This can be achieved by conducting row permutations evenly over the entire length of the simulated series.

To achieve computational efficiency, each permutation is evaluated in terms of its impact on the value of function D. If the value becomes closer to ε, the permutation is accepted. If not, the permutation is rejected and a new possible permutation is inspected. Again, the sums for all the above statistics are not re-calculated every time, but are merely updated to account for the changes caused by the two swapped rows. The proposed algorithm conducts efficient deterministic search in a way that closely resembles the bubble sort algorithm. The basic ideas are summarized in the pseudo code below:

```
set the initial value of D to Di
for rows i = 1, 1000  // start from the top
  for rows j = 1000, 1 // start from the bottom
    if( i ≠ j )
      swap rows (i, j)
      update D
      if(D < Di){
        accept the swap
        set Di = D}
      if( D < ε )
        exit
  next  j
next  i
```

One could argue that the above algorithm cannot guarantee convergence below the desired threshold ε. Indeed, for a small sample of say only 100 synthetic years, there may be not enough combinations available to achieve the desired accuracy. In that case the algorithm will find a combination with the closest value of D to ε. However, it is always possible to generate sufficient number of years to provide a pool of data with enough possible permutations to achieve convergence, even if only a partial subset of the generated series may be required in the end. In the numerical tests conducted so far it was found that a pool of 1,000 years of synthetic data is more than sufficient to handle the most stringent requirements. If for example only 200 years of synthetic data are required for a given study, one could pick any subset of 200 sequential years from the generated pool of 1,000 years.

5 Numerical example

Four historical flow series from southern Alberta in western Canada with naturalized weekly flows were extracted from Alberta Environment's natural flow database. The available historical records that were used in the analyses were from 1912 to 1995 on all four stations. The stations are listed in Table 3 and their configuration in Fig. 1. Three stations are located on the Belly River, and one station is located on the Waterton River which is

Table 3 Selected stations

Station	Code	Station name
1	GBWCON	Belly river below confluence with Waterton river
2	C5AD41	Belly river near Glenwood
3	C5AD05	Belly river at mountain view
4	C5AD26	Waterton river at Waterton reservoir

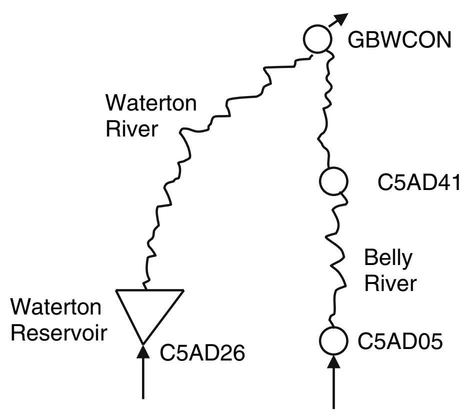

Fig. 1 Locations of selected stations relative to each other

Table 4 Summary of annual statistics

	Station 1	Station 2	Station 3	Station 4
Summary of simulated annual statistics				
Mean	34.342	9.511	9.069	22.246
St. dev	9.682	2.302	2.018	5.987
Skew	0.969	0.890	0.832	0.905
Summary of historical annual statistics				
Mean	34.336	9.510	9.069	22.247
St. Dev	9.796	2.305	1.995	6.030
Skew	1.092	0.693	0.369	0.838
Cross-correlation matrix of simulated annual flows				
Station 1	1.000	0.962	0.936	0.985
Station 2		1.000	0.977	0.943
Station 3			1.000	0.944
Station 4				1.000
Cross-correlation matrix of historical annual flows				
Station 1	1.000	0.966	0.942	0.988
Station 2		1.000	0.980	0.946
Station 3			1.000	0.945
Station 4				1.000

its tributary. With this configuration, it is reasonable to expect a high degree of both auto-correlation and cross-correlation between all four stations. As in the previous work of this nature, a comparison between the historical and simulated statistics is given in Tables 4 through 8. Although the target order of weekly autoregressive relationship of crossing from year i to year i +1 was only lag 1 (i.e. function D_k in Eq. 2 was evaluated only on the bases of correlation between week 52 and week 1 of the subsequent year), it can be noticed that correlations of lag 2 also match well by inference, due to a good match of correlations between weeks 51 and 52 as well as between weeks 1 and 2 which was achieved in step 2 of the proposed algorithm. Similarly, the annual statistics in Table 4 show a good match between the historical and simulated annual series although they were not directly handled by the algorithm, but were rather a consequence of a good match previously achieved with the weekly statistics shown in Table 5. Finally, reordering of the generated years of data was conducted to induce a match of annual auto correlations. The model was run to match 15 lags of the annual auto correlation and the results are shown in Fig. 2. Although there is not much annual auto correlation worth preserving here, the algorithm demonstrates that it is capable of finding the sequence of generated years of data that can match the desired target autocorrelations for all 4 stations and for all 15 lags simultaneously.

In addition to the summary and comparison of the target statistics of the simulated and historical series, it is useful to validate the generated series on the basis of its extreme low and high flows. After all, the goal of stochastic hydrology is to produce a possible realization of continuous stream flows that contain extreme events that had not been seen in the much shorter historical record. Without stochastic hydrology, such extreme events are typically assessed using the frequency analyses approach (which provides only the peak flow estimates), where an annual series of extreme low or high flows is input into a model in an effort to fit the data to one of the theoretical distributions popular in hydrology. In this example, the 84 years of historical annual minimum and maximum weekly flows were input into a frequency analyses program to obtain the estimates of weekly flows with return periods of 100, 200 and 500 years, for both low and high flow events. As usual, the estimates of extreme flow events were obtained by relying on the quality of the functional fit provided using the maximum likelihood method as well as on the length of the input data record, and it is understood that a 500 year flood cannot be

Table 5 Comparison of historical and simulated weekly averages and standard deviations

Week	Station 1 (GBWCON)				Station 2 (C5AD41)				Station 3 (C5AD05)				Station 4 (C5AD26)			
	Simulated Average	Historic Average	Simulated St. Dev	Historic St. Dev	Simulated Average	Historic Average	Simulated St. Dev	Historic St. Dev	Simulated Average	Historic Average	Simulated St. Dev	Historic St. Dev	Simulated Average	Historic Average	Simulated St. Dev	Historic St. Dev
1	7.386	7.388	4.239	4.645	2.125	2.126	1.176	1.292	2.019	2.021	1.098	1.177	4.797	4.796	2.726	3.020
2	7.090	7.092	3.227	3.315	2.025	2.024	0.858	0.883	1.895	1.894	0.779	0.828	4.598	4.596	2.311	2.260
3	7.274	7.271	3.675	3.684	2.047	2.042	0.988	0.961	1.855	1.854	0.840	0.827	4.697	4.696	2.284	2.323
4	6.864	6.873	3.103	3.092	1.991	1.989	0.966	0.947	1.822	1.822	0.879	0.875	4.455	4.451	2.046	2.057
5	6.953	6.941	3.161	3.082	1.996	1.997	0.918	0.922	1.785	1.785	0.840	0.840	4.424	4.426	1.960	1.934
6	6.880	6.874	3.143	2.985	1.988	1.988	0.952	0.947	1.819	1.818	0.898	0.908	4.303	4.299	1.784	1.813
7	7.744	7.724	5.497	6.204	2.228	2.233	1.974	2.058	2.002	2.006	1.852	1.907	4.676	4.666	2.793	2.982
8	7.702	7.704	4.770	5.273	2.141	2.142	1.138	1.166	1.851	1.853	0.868	0.839	4.665	4.670	2.656	2.840
9	8.608	8.650	6.846	7.533	2.485	2.448	2.505	2.420	2.068	2.072	2.306	2.268	5.160	5.154	4.054	4.329
10	8.552	8.568	4.582	4.730	2.339	2.338	1.319	1.373	1.877	1.877	0.954	0.993	5.078	5.080	2.737	2.754
11	10.280	10.288	9.458	9.607	2.664	2.652	2.218	2.400	2.020	2.019	1.163	1.183	5.780	5.781	3.771	4.002
12	12.551	12.497	9.186	9.429	3.079	3.077	2.063	2.219	2.151	2.153	1.302	1.293	6.789	6.803	4.474	4.385
13	14.815	14.806	9.730	10.119	3.519	3.520	2.191	2.173	2.426	2.423	1.393	1.349	8.050	8.045	5.294	4.771
14	17.727	17.710	12.909	12.503	4.159	4.160	2.374	2.196	3.208	3.211	1.939	1.824	10.045	10.047	6.971	6.767
15	20.348	20.417	11.146	10.715	5.313	5.319	2.907	2.864	4.503	4.502	2.907	2.876	12.003	11.996	6.745	6.404
16	27.669	27.615	15.868	15.517	7.683	7.685	4.357	4.252	6.942	6.945	4.424	4.334	16.991	16.977	10.798	10.186
17	37.828	37.834	22.092	20.826	10.441	10.440	6.385	5.962	9.605	9.608	6.234	6.074	24.090	24.042	14.495	13.934
18	51.290	51.287	27.445	25.815	13.846	13.861	7.603	7.014	13.080	13.067	7.249	6.890	33.479	33.512	18.206	17.157
19	71.186	71.141	32.125	30.402	18.456	18.457	8.677	8.034	17.587	17.612	8.102	7.643	47.418	47.451	22.086	20.534
20	95.488	95.522	39.228	35.233	24.532	24.538	9.692	8.452	23.776	23.737	9.284	8.148	64.418	64.459	25.155	23.693
21	117.599	117.625	52.728	49.113	29.662	29.635	12.707	11.822	29.025	29.006	12.276	11.162	80.797	80.781	34.773	33.133
22	124.505	124.615	54.963	49.457	30.786	30.829	12.204	10.924	30.186	30.202	11.229	10.219	86.198	86.265	36.253	33.454
23	139.880	139.834	71.725	77.365	35.112	35.138	19.205	20.029	34.082	34.010	16.399	16.750	95.726	95.745	48.442	47.235
24	134.072	134.121	63.286	62.158	33.272	33.299	14.720	14.348	32.597	32.606	14.288	13.449	92.482	92.452	42.204	42.111
25	125.872	125.804	73.030	75.055	31.303	31.280	16.668	17.201	30.906	30.894	17.093	16.802	84.836	84.868	47.181	47.458
26	104.051	103.837	57.499	52.872	27.106	27.123	12.808	11.672	26.371	26.404	11.765	10.400	69.083	68.977	35.135	32.466
27	82.979	82.954	45.542	42.306	23.468	23.442	10.796	10.246	22.513	22.537	10.090	9.465	54.243	54.263	30.044	28.314
28	65.533	65.511	34.239	33.420	19.750	19.721	9.443	8.510	19.082	19.077	8.664	8.218	42.118	42.110	24.113	22.754
29	51.849	51.698	26.134	25.454	16.120	16.106	7.160	6.475	15.573	15.582	6.895	6.164	32.477	32.414	17.803	17.131
30	41.463	41.445	21.781	20.649	13.350	13.340	5.489	4.992	12.880	12.888	5.112	4.626	25.232	25.226	13.838	13.145
31	31.811	31.832	14.334	13.671	10.962	10.958	4.109	3.691	10.592	10.593	3.779	3.358	18.829	18.849	9.284	8.563
32	25.752	25.735	11.100	11.036	9.115	9.106	3.384	3.059	8.877	8.873	3.054	2.831	14.954	14.950	6.992	6.847
33	22.537	22.552	10.448	9.968	8.212	8.213	2.943	2.858	7.861	7.864	2.758	2.566	12.767	12.778	6.311	6.064

Table 5 continued

Week	Station 1 (GBWCON)				Station 2 (C5AD41)				Station 3 (C5AD05)				Station 4 (C5AD26)			
	Simulated Average	Historic Average	Simulated St. Dev	Historic St. Dev	Simulated Average	Historic Average	Simulated St. Dev	Historic St. Dev	Simulated Average	Historic Average	Simulated St. Dev	Historic St. Dev	Simulated Average	Historic Average	Simulated St. Dev	Historic St. Dev
34	20.316	20.311	9.464	9.179	7.456	7.454	2.833	2.724	7.074	7.072	2.533	2.400	11.304	11.302	5.359	5.304
35	18.998	19.004	11.936	13.073	6.670	6.660	2.838	2.955	6.355	6.349	2.615	2.589	10.799	10.806	7.377	7.962
36	17.634	17.642	12.125	12.662	6.198	6.200	2.899	2.800	5.955	5.947	2.615	2.545	10.077	10.067	7.697	8.109
37	17.885	17.877	14.684	15.339	6.057	6.055	3.337	3.599	5.700	5.699	2.956	3.074	10.195	10.269	8.177	9.308
38	17.372	17.416	12.432	12.980	5.777	5.774	3.072	2.996	5.427	5.426	2.689	2.494	10.227	10.193	8.176	8.439
39	18.153	18.117	15.201	15.807	5.750	5.770	3.260	3.554	5.423	5.425	3.000	3.034	10.558	10.641	8.976	9.686
40	17.380	17.432	14.836	15.574	5.533	5.526	2.896	3.325	5.264	5.264	2.766	3.010	10.422	10.407	10.460	10.361
41	17.379	17.389	12.281	12.415	5.676	5.676	3.358	3.073	5.365	5.357	3.048	2.847	10.181	10.177	7.871	8.089
42	17.953	17.972	14.186	13.465	5.947	5.949	4.425	4.478	5.646	5.646	4.339	4.251	10.734	10.720	8.767	8.420
43	17.599	17.596	13.568	12.799	5.368	5.373	3.303	3.165	5.131	5.134	3.163	3.077	10.823	10.843	8.812	8.639
44	15.614	15.565	10.684	10.471	4.868	4.871	2.680	2.721	4.687	4.689	2.649	2.681	9.627	9.636	7.278	7.103
45	14.555	14.545	8.650	8.916	4.571	4.573	2.876	3.048	4.444	4.441	2.918	3.153	9.140	9.154	5.631	5.774
46	14.285	14.289	9.286	9.693	4.265	4.260	2.891	3.028	4.049	4.047	2.456	2.719	9.128	9.119	6.430	6.368
47	12.201	12.229	6.576	7.180	3.576	3.578	1.926	2.043	3.370	3.368	1.715	1.811	7.893	7.884	4.675	4.934
48	11.342	11.361	6.328	6.478	3.359	3.362	1.964	2.021	3.179	3.177	1.950	2.130	7.277	7.270	3.992	4.252
49	9.917	9.915	5.921	5.876	2.924	2.922	1.748	1.959	2.731	2.728	1.745	1.916	6.435	6.430	3.728	3.879
50	9.084	9.075	4.512	4.607	2.685	2.684	1.425	1.512	2.529	2.535	1.404	1.468	5.838	5.834	2.993	2.972
51	8.279	8.272	3.618	3.644	2.411	2.411	1.029	1.001	2.343	2.346	0.995	0.961	5.414	5.403	2.566	2.512
52	7.703	7.700	3.677	3.779	2.216	2.211	1.007	0.972	2.105	2.104	0.915	0.910	5.062	5.075	2.411	2.652

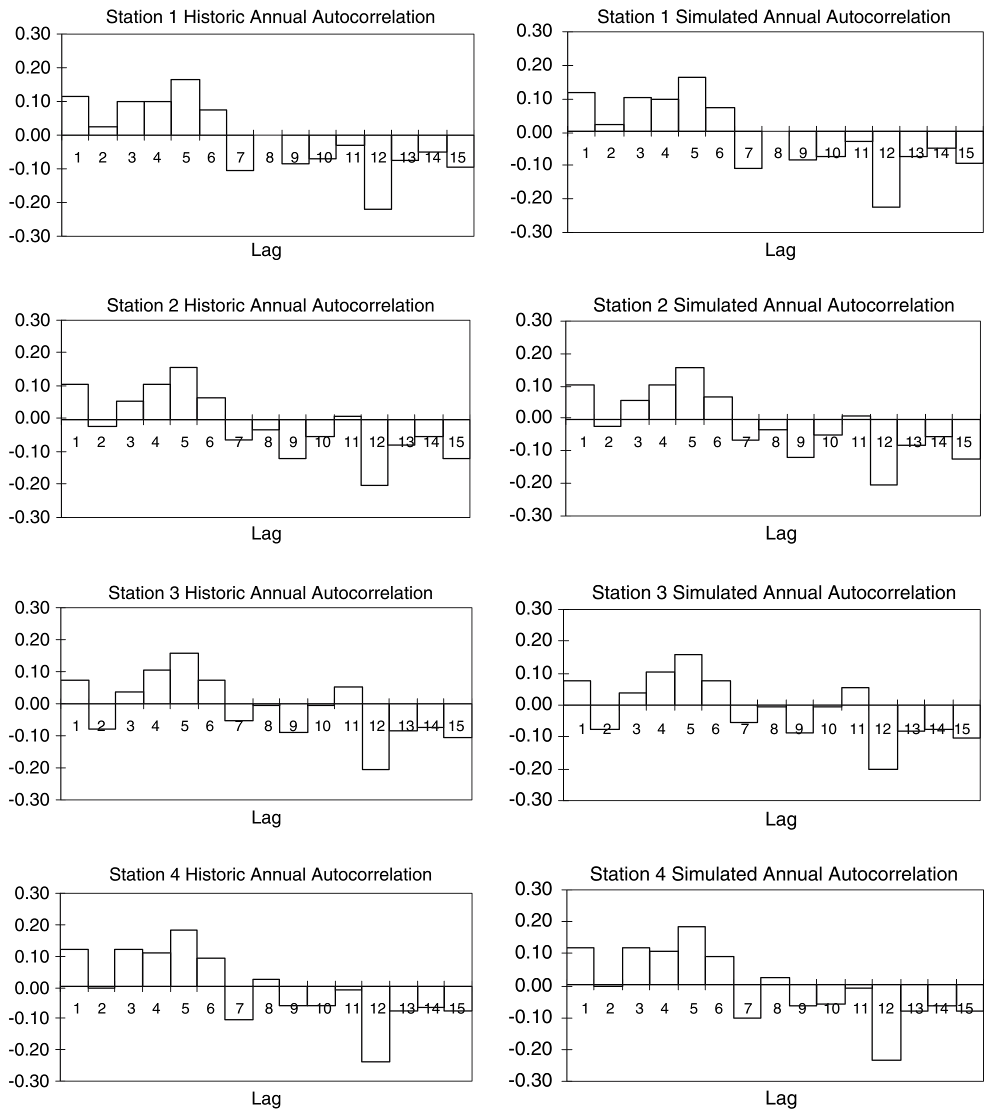

Fig. 2 Comparison of the historic and simulated annual auto correlations

estimated with much certainty based on 84 years of data. Also, the statistical distributions used for flood frequency analyses are not particularly suitable for weekly flows, but were rather aimed to fit daily or instantaneous flows. However, it was felt that the values obtained from the frequency analyses should be comparable to the flows corresponding to the same return periods in the synthetic series. The two best fits that were obtained with the available historical datasets were based on the 3 parameter Log Normal and Gumbel III statistical distributions. These were then compared with high flows and low flows obtained from the series of 1,000 years of annual minimum and maximum weekly flows available from the synthetically generated series, using the Weibull plotting position formula of the form rank/(n + 1). Table 8 shows a reasonably good agreement between the extreme flow events on all four stations obtained from the frequency analyses, with the readings from the sorted sequences of synthetic annual minimum and maximum weekly flows that have the probabilities of occurrence corresponding to the return flow periods of 100, 200 and 500 years.

6 Conclusions

This paper presents a new approach to generating stochastic flow series. It holds out a promise to offer a practical and user friendly tool to the practitioners. The proposed method demonstrates ability to provide a close match between the relevant statistics of the synthetic and historical hydrologic series. Research efforts are currently under way to extend the ideas presented here to address simulation of daily flows as well as dry region flows, which may have multiple time periods with zero values. The results of these efforts will be addressed in future publications.

Table 6 Historical weekly cross-correlations and auto-correlations of Lag 1 and Lag 2

Week	Cross correlations between Stations 1, 2, 3 and 4						Station 1		Station 2		Station 3		Station 4	
	1–2	1–3	1–4	2–3	2–4	3–4	Lag 1	Lag 2	Lag 1	Lag 2	Lag 1	Lag 2	Lag 1	Lag 2
1	0.952	0.942	0.987	0.983	0.907	0.914	0.837	0.491	0.734	0.487	0.772	0.587	0.877	0.544
2	0.913	0.849	0.957	0.927	0.797	0.802	0.625	0.579	0.691	0.496	0.749	0.476	0.653	0.580
3	0.905	0.780	0.934	0.845	0.772	0.808	0.854	0.577	0.749	0.506	0.671	0.416	0.893	0.647
4	0.893	0.860	0.952	0.939	0.764	0.805	0.784	0.627	0.848	0.560	0.844	0.589	0.818	0.722
5	0.921	0.840	0.951	0.933	0.821	0.794	0.806	0.365	0.746	0.256	0.778	0.230	0.858	0.462
6	0.906	0.821	0.968	0.950	0.838	0.790	0.521	0.499	0.362	0.462	0.288	0.467	0.627	0.545
7	0.975	0.917	0.979	0.975	0.944	0.903	0.865	0.329	0.702	0.175	0.481	0.145	0.879	0.388
8	0.937	0.776	0.985	0.921	0.920	0.796	0.579	0.605	0.585	0.463	0.637	0.631	0.605	0.674
9	0.984	0.932	0.992	0.958	0.976	0.954	0.708	0.292	0.698	0.278	0.696	0.384	0.746	0.425
10	0.891	0.757	0.960	0.848	0.831	0.798	0.564	0.314	0.644	0.376	0.743	0.604	0.656	0.387
11	0.978	0.768	0.955	0.805	0.915	0.850	0.548	0.249	0.598	0.263	0.802	0.476	0.680	0.407
12	0.945	0.729	0.963	0.827	0.887	0.791	0.570	0.314	0.497	0.394	0.606	0.458	0.713	0.407
13	0.916	0.455	0.933	0.618	0.806	0.605	0.810	0.629	0.797	0.499	0.776	0.528	0.724	0.611
14	0.856	0.405	0.976	0.734	0.834	0.505	0.735	0.502	0.663	0.360	0.748	0.468	0.752	0.561
15	0.898	0.754	0.977	0.937	0.906	0.825	0.725	0.449	0.705	0.410	0.721	0.483	0.760	0.526
16	0.891	0.815	0.988	0.969	0.891	0.844	0.706	0.407	0.662	0.350	0.670	0.366	0.733	0.425
17	0.935	0.884	0.985	0.974	0.925	0.906	0.624	0.346	0.597	0.255	0.643	0.323	0.664	0.385
18	0.928	0.863	0.986	0.974	0.926	0.891	0.514	0.145	0.427	0.029	0.500	0.021	0.568	0.129
19	0.940	0.876	0.983	0.930	0.926	0.918	0.517	0.054	0.404	–0.034	0.338	–0.090	0.518	0.016
20	0.904	0.842	0.987	0.967	0.889	0.859	0.523	0.156	0.438	0.127	0.368	0.120	0.489	0.132
21	0.962	0.920	0.993	0.971	0.957	0.937	0.606	0.313	0.613	0.290	0.603	0.269	0.595	0.320
22	0.949	0.904	0.987	0.970	0.936	0.920	0.613	0.442	0.528	0.389	0.507	0.349	0.597	0.414
23	0.965	0.883	0.985	0.945	0.948	0.917	0.653	0.519	0.718	0.422	0.727	0.430	0.657	0.374
24	0.955	0.944	0.992	0.986	0.936	0.940	0.583	0.498	0.602	0.563	0.616	0.595	0.592	0.516
25	0.950	0.947	0.940	0.996	0.975	0.974	0.720	0.684	0.720	0.751	0.751	0.785	0.716	0.693
26	0.967	0.931	0.961	0.978	0.973	0.967	0.861	0.779	0.878	0.786	0.871	0.782	0.877	0.800
27	0.969	0.939	0.995	0.981	0.962	0.948	0.900	0.797	0.906	0.827	0.920	0.830	0.917	0.804
28	0.972	0.957	0.996	0.989	0.965	0.959	0.917	0.674	0.906	0.713	0.897	0.728	0.920	0.686
29	0.961	0.958	0.995	0.986	0.943	0.950	0.829	0.798	0.855	0.806	0.874	0.805	0.841	0.793
30	0.951	0.938	0.994	0.990	0.940	0.938	0.880	0.748	0.884	0.772	0.886	0.810	0.888	0.766
31	0.938	0.931	0.991	0.982	0.901	0.909	0.863	0.771	0.857	0.726	0.864	0.722	0.881	0.780
32	0.905	0.902	0.991	0.980	0.857	0.871	0.877	0.683	0.802	0.611	0.847	0.638	0.903	0.728
33	0.890	0.851	0.988	0.962	0.830	0.811	0.803	0.627	0.772	0.583	0.791	0.646	0.843	0.651
34	0.908	0.859	0.968	0.953	0.819	0.825	0.703	0.666	0.710	0.616	0.798	0.608	0.748	0.667
35	0.877	0.814	0.995	0.978	0.852	0.798	0.926	0.606	0.780	0.523	0.804	0.552	0.946	0.651
36	0.852	0.803	0.991	0.971	0.800	0.765	0.698	0.683	0.703	0.637	0.698	0.610	0.730	0.664
37	0.926	0.911	0.993	0.976	0.889	0.885	0.936	0.703	0.848	0.608	0.811	0.568	0.948	0.724
38	0.901	0.875	0.994	0.971	0.861	0.844	0.842	0.771	0.827	0.691	0.833	0.664	0.832	0.746
39	0.968	0.954	0.997	0.987	0.958	0.951	0.951	0.856	0.893	0.549	0.870	0.487	0.949	0.889
40	0.947	0.944	0.996	0.984	0.923	0.928	0.908	0.656	0.684	0.461	0.617	0.441	0.935	0.674
41	0.851	0.800	0.989	0.968	0.778	0.733	0.840	0.682	0.705	0.618	0.698	0.594	0.843	0.663
42	0.931	0.913	0.995	0.994	0.899	0.882	0.885	0.583	0.847	0.583	0.830	0.568	0.879	0.592
43	0.945	0.927	0.995	0.988	0.927	0.914	0.755	0.598	0.829	0.492	0.829	0.461	0.750	0.619
44	0.952	0.932	0.993	0.982	0.929	0.919	0.750	0.448	0.576	0.363	0.547	0.405	0.775	0.502
45	0.920	0.885	0.991	0.987	0.880	0.850	0.613	0.554	0.566	0.506	0.658	0.579	0.649	0.577
46	0.965	0.957	0.995	0.988	0.948	0.948	0.875	0.754	0.808	0.654	0.855	0.599	0.898	0.791
47	0.938	0.921	0.989	0.932	0.901	0.919	0.905	0.736	0.878	0.694	0.859	0.688	0.919	0.766

Table 6 continued

Week	Cross correlations between Stations 1, 2, 3 and 4						Station 1		Station 2		Station 3		Station 4	
	1–2	1–3	1–4	2–3	2–4	3–4	Lag 1	Lag 2	Lag 1	Lag 2	Lag 1	Lag 2	Lag 1	Lag 2
48	0.917	0.877	0.990	0.947	0.885	0.864	0.889	0.801	0.814	0.721	0.804	0.710	0.904	0.802
49	0.912	0.883	0.985	0.971	0.861	0.852	0.871	0.708	0.878	0.727	0.888	0.723	0.887	0.728
50	0.913	0.862	0.983	0.972	0.860	0.831	0.882	0.652	0.843	0.594	0.794	0.554	0.890	0.673
51	0.917	0.814	0.971	0.942	0.850	0.773	0.814	0.580	0.791	0.601	0.781	0.565	0.803	0.536
52	0.879	0.834	0.974	0.924	0.804	0.801	0.832	0.767	0.729	0.691	0.763	0.734	0.861	0.754

Table 7 Simulated weekly cross-correlations and auto-correlations of Lag 1 and Lag 2

Week	Cross correlations between Stations 1, 2, 3 and 4						Station 1		Station 2		Station 3		Station 4	
	1–2	1–3	1–4	2–3	2–4	3–4	Lag 1	Lag 2	Lag 1	Lag 2	Lag 1	Lag 2	Lag 1	Lag 2
1	0.946	0.932	0.979	0.976	0.899	0.906	0.809	0.537	0.745	0.493	0.765	0.611	0.869	0.573
2	0.912	0.845	0.935	0.929	0.811	0.803	0.634	0.578	0.677	0.477	0.741	0.437	0.669	0.607
3	0.885	0.771	0.921	0.827	0.747	0.800	0.849	0.521	0.743	0.455	0.611	0.360	0.889	0.605
4	0.882	0.816	0.933	0.912	0.745	0.739	0.742	0.583	0.808	0.535	0.805	0.562	0.783	0.694
5	0.894	0.811	0.934	0.903	0.777	0.753	0.790	0.320	0.724	0.210	0.762	0.181	0.861	0.463
6	0.881	0.786	0.958	0.935	0.821	0.766	0.456	0.506	0.284	0.427	0.205	0.425	0.604	0.515
7	0.930	0.814	0.955	0.952	0.878	0.779	0.932	0.258	0.577	0.069	0.336	0.016	0.834	0.247
8	0.826	0.625	0.895	0.895	0.892	0.747	0.408	0.493	0.397	0.390	0.411	0.547	0.460	0.614
9	0.871	0.747	0.938	0.819	0.911	0.834	0.682	0.168	0.449	0.059	0.422	0.149	0.590	0.238
10	0.859	0.658	0.960	0.705	0.813	0.695	0.455	0.227	0.416	0.245	0.672	0.509	0.510	0.274
11	0.877	0.643	0.941	0.641	0.835	0.754	0.483	0.189	0.521	0.228	0.743	0.414	0.579	0.307
12	0.943	0.657	0.942	0.743	0.867	0.700	0.482	0.282	0.457	0.296	0.491	0.355	0.608	0.322
13	0.936	0.366	0.922	0.453	0.832	0.506	0.864	0.542	0.702	0.406	0.725	0.432	0.638	0.523
14	0.764	0.305	0.891	0.694	0.781	0.441	0.638	0.441	0.618	0.344	0.680	0.420	0.696	0.533
15	0.879	0.713	0.966	0.918	0.880	0.785	0.687	0.394	0.673	0.321	0.675	0.414	0.708	0.471
16	0.867	0.780	0.976	0.965	0.859	0.799	0.635	0.327	0.580	0.262	0.588	0.266	0.648	0.311
17	0.899	0.846	0.975	0.962	0.877	0.862	0.564	0.325	0.530	0.230	0.567	0.276	0.584	0.355
18	0.921	0.844	0.980	0.964	0.923	0.878	0.502	0.129	0.403	0.022	0.478	0.005	0.557	0.110
19	0.918	0.833	0.972	0.903	0.903	0.883	0.464	0.048	0.325	–0.043	0.228	–0.103	0.446	0.010
20	0.868	0.790	0.975	0.947	0.846	0.809	0.497	0.130	0.412	0.095	0.349	0.082	0.473	0.119
21	0.947	0.908	0.987	0.957	0.941	0.926	0.567	0.249	0.597	0.204	0.586	0.205	0.572	0.241
22	0.931	0.877	0.976	0.956	0.916	0.898	0.568	0.401	0.462	0.383	0.451	0.329	0.520	0.375
23	0.929	0.845	0.973	0.923	0.921	0.891	0.574	0.497	0.637	0.361	0.660	0.357	0.590	0.295
24	0.943	0.930	0.983	0.974	0.918	0.926	0.560	0.439	0.532	0.522	0.546	0.562	0.525	0.485
25	0.917	0.909	0.895	0.986	0.950	0.950	0.661	0.616	0.667	0.686	0.675	0.694	0.662	0.637
26	0.950	0.912	0.947	0.966	0.964	0.960	0.836	0.753	0.864	0.766	0.854	0.769	0.862	0.787
27	0.959	0.930	0.992	0.975	0.954	0.941	0.891	0.782	0.899	0.812	0.916	0.816	0.912	0.788
28	0.964	0.949	0.993	0.984	0.954	0.951	0.901	0.649	0.892	0.687	0.879	0.704	0.901	0.664
29	0.952	0.949	0.988	0.981	0.926	0.935	0.813	0.781	0.845	0.800	0.864	0.796	0.817	0.775
30	0.937	0.925	0.989	0.986	0.930	0.929	0.876	0.718	0.886	0.764	0.881	0.806	0.887	0.746
31	0.925	0.920	0.986	0.975	0.890	0.900	0.841	0.749	0.842	0.709	0.854	0.696	0.864	0.750
32	0.887	0.883	0.981	0.972	0.831	0.845	0.838	0.619	0.779	0.576	0.829	0.606	0.866	0.656
33	0.852	0.822	0.981	0.939	0.778	0.776	0.776	0.601	0.759	0.516	0.767	0.605	0.807	0.548
34	0.900	0.852	0.953	0.947	0.799	0.810	0.702	0.605	0.688	0.585	0.781	0.565	0.675	0.570
35	0.747	0.677	0.918	0.963	0.792	0.721	0.866	0.586	0.740	0.484	0.772	0.526	0.913	0.577
36	0.691	0.641	0.956	0.956	0.731	0.680	0.617	0.570	0.665	0.610	0.659	0.576	0.627	0.570

Table 7 continued

Week	Cross correlations between Stations 1, 2, 3 and 4						Station 1		Station 2		Station 3		Station 4	
	1–2	1–3	1–4	2–3	2–4	3–4	Lag 1	Lag 2	Lag 1	Lag 2	Lag 1	Lag 2	Lag 1	Lag 2
37	0.762	0.753	0.901	0.974	0.840	0.842	0.892	0.664	0.797	0.532	0.778	0.506	0.916	0.618
38	0.773	0.749	0.959	0.968	0.834	0.822	0.749	0.624	0.808	0.696	0.813	0.673	0.766	0.651
39	0.916	0.886	0.920	0.983	0.944	0.939	0.888	0.875	0.880	0.546	0.860	0.478	0.864	0.863
40	0.889	0.901	0.942	0.967	0.833	0.871	0.894	0.668	0.652	0.448	0.604	0.453	0.845	0.602
41	0.738	0.680	0.957	0.967	0.764	0.714	0.804	0.603	0.703	0.611	0.691	0.591	0.827	0.665
42	0.816	0.782	0.951	0.982	0.886	0.862	0.811	0.549	0.829	0.599	0.806	0.570	0.870	0.612
43	0.927	0.911	0.990	0.980	0.915	0.906	0.752	0.564	0.831	0.411	0.818	0.353	0.754	0.579
44	0.941	0.930	0.986	0.974	0.922	0.920	0.717	0.440	0.494	0.303	0.437	0.323	0.736	0.439
45	0.860	0.796	0.978	0.966	0.803	0.745	0.606	0.531	0.488	0.464	0.585	0.523	0.585	0.555
46	0.907	0.904	0.966	0.962	0.912	0.906	0.810	0.729	0.730	0.553	0.794	0.488	0.861	0.741
47	0.915	0.879	0.980	0.894	0.876	0.883	0.914	0.729	0.831	0.632	0.810	0.584	0.893	0.727
48	0.865	0.784	0.946	0.913	0.862	0.808	0.854	0.728	0.749	0.660	0.721	0.605	0.881	0.770
49	0.838	0.770	0.937	0.927	0.787	0.755	0.838	0.648	0.795	0.626	0.776	0.596	0.848	0.643
50	0.870	0.772	0.971	0.935	0.800	0.731	0.835	0.629	0.781	0.550	0.713	0.480	0.846	0.607
51	0.887	0.787	0.959	0.936	0.819	0.744	0.836	0.670	0.759	0.626	0.731	0.583	0.773	0.646
52	0.818	0.763	0.934	0.906	0.770	0.762	0.832	0.686	0.727	0.545	0.761	0.563	0.859	0.742

Table 8 Comparison of extreme flow estimates

	Low flow return period (years)			High flow return period (years)		
	500	200	100	100	200	500
Station 1						
3 Par. Lognormal	1.72	1.92	2.09	467.64	522.81	598.56
Gumbel type III	1.95	2.05	2.15	488.03	555.98	653.50
Plotting position formula	1.69	1.79	1.88	436.01	491.08	552.95
Station 2						
3 Par. Lognormal	0.34	0.43	0.50	111.43	124.55	142.64
Gumbel type III	0.43	0.48	0.53	118.91	136.73	163.04
Plotting position formula	0.41	0.42	0.46	115.29	146.83	169.78
Station 3						
3 Par. Lognormal	0.32	0.39	0.45	99.75	110.09	124.10
Gumbel type III	0.39	0.43	0.47	102.29	114.16	130.63
Plotting position formula	0.32	0.34	0.38	104.49	115.31	146.20
Station 4						
3 Par. Lognormal	0.90	1.06	1.19	295.57	325.76	366.32
Gumbel type III	1.11	1.18	1.26	299.15	331.15	374.21
Plotting position formula	1.13	1.19	1.23	308.35	329.78	353.31

Acknowledgments The authors wish to express appreciation to the National Scientific and Engineering Research Council (NSERC) of Canada, and to Golder Associates Ltd. of Calgary, Canada, for jointly sponsoring this research project. The acknowledgment is also extended to Dr. Shuhong Wu of Golder Associates Ltd. for providing his program for low flow and high flow frequency analyses that helped obtain the statistics necessary for compiling information presented in Table 8.

References

Askoy H, Bayazit M (2000) A model for daily flows of intermittent streams. Hydrol Processes 14:1725–1744

Box GEP, Jenkins G (1970) Time series analysis, forecasting and control, 1st edn. Holden Day, San Francisco

Cario MC, Nelson BL (1996) Autoregressive to anything: time series input processes for simulation. Oper Res Lett 19:51–58

Cario MC, Nelson BL (1997) Numerical methods for fitting and simulation autoregressive-to-anything processes. INFORMS J Comput 10:72–81

Cooley WW, Lohnes PR (1971) Multivariate data analysis, Chap 3. Robert E. Krieger Publishing Company, Malabar, Florida

Devore L (1986) Non-uniform random variate generation. Springer, Heidelberg

Ghosh S, Henderson SG (2002) Chessboard distributions and random vectors with specified marginals and covariance matrix. Oper Res 50(5):820–834

Hurst H (1957) A suggested statistical model for some time series that occur in nature. Nature 180:494

Ilich N (2006) A matching algorithm for generating statistically dependent variables with arbitrary marginals. Euro J Oper Res (to appear)

Iman R, Conover W (1982) A distribution free approach to inducing a rank correlation among input variables. Commun Stat Simul Comput 11(3):311–334

Kendall DR, Dracup JA (1991) A comparison of index-sequential and AR(1) generated hydrologic sequences. J Hydrol 122(1–4):335–352

Lall U (1995) Recent advances in nonparametric function estimation: hydrology applications. Review of Geophysics, Supplement, July 1995. U.S. National Report to International Union of Geodesy and Geophysics 1991–1994, pp 1093–1102

Matalas NC (1967) Mathematical assessment of synthetic hydrology. Water Resources Res 3(4):937–945

Mejia JM, Rousselle J (1976) Disaggregation models in hydrology revisited. Water Resources Res 12(2):185–186

Moon YU, Lall U, Bosworth K (1993) A comparison of tail probability estimators for flood frequency analyses. J Hydrol 151:343–363

Rasmussen PF, Salas JD, Fagherazzi L, Rassam JC, Bobbe B (1996) Estimation and validation of contemporaneous PARMA models for streamflow simulation. Water Resources Res 32(10):3151–3160

Sharma A, Tarboton DG, Lall U (1997) Streamflow Simulation: a nonparametric approach. Water Resources Res 33(2):291–308

Srinivas VV, Srinivasan K (2005) Hybrid moving block bootstrap for stochastic simulation of multi-site streamflows. J Hydrol 302:307–330

Sveinsson OGB, Salas JD (2003) Modeling of dynamics of long-term variability of hydroclimatic processes. J Hydrometeorol 4(3):489–505

Thomas HA Jr, Fiering MB (1962) Mathematical synthesis of streamflow sequences for analyses of river basins by simulation. In: Maas A et al (eds) The design of water resources system. Harvard University Press, Cambridge, pp 459–493

UNESCO (2004) IDAMS: internationally developed data analysis and management software package – WinIDAMS Reference Manual, Release 1.2. http://www.unesco.org/idams

US Army Corps of Engineers (1971) HEC-4 User Manual

Valencia DR, Schaake JC (1973) Disaggregation processes in stochastic hydrology. Water Resources Res 9(3):580–585